

A data storage system for automatically providing and maintaining secondary data on a secondary data storage device (50b) which is generally identical copy of primary data stored on a primary data storage device (22a) is presented. In particular, the secondary data storage device (50b) on which is maintained a copy of the primary data is geographically physically remote from the primary data storage device (22a) while the primary data storage system controller (16), coupled to the secondary data storage device (50b) by high speed communication link (40), controls the copying and maintaining a list of primary data which is to be copied to the secondary data storage device (50b) without intervention of a primary host computer (12).

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

REMOTE DATA MIRRORINGRELATED APPLICATION

This Application is a continuation in part of U.S. Patent Application Nos. 07/586,796 filed September 24, 1990 entitled SYSTEM AND METHOD FOR DISK MAPPING AND DATA RETRIEVAL; 07/587,247  
5 filed September 24, 1990 entitled DYNAMICALLY RECONFIGURABLE DATA STORAGE SYSTEM and 07/587,253 filed September 24, 1990 entitled RECONFIGURABLE MULTI-FUNCTION DISK CONTROLLER which are fully incorporated herein by reference.

FIELD OF THE INVENTION

10 This invention relates to data storage on disk drives and more particularly, to a system and method for automatically providing and maintaining a copy or mirror of a data storage disk at a location remote from the main data storage disk.

BACKGROUND OF THE INVENTION

15 Nearly all data processing system users are concerned with maintaining back up data in order to insure continued data processing operations should their data become lost, damaged, or unavailable.

20 Large institutional users of data processing systems which maintain large volumes of data such as banks, insurance companies, and stock market traders must and do take tremendous steps to insured back up data availability in case of a major disaster.

25 These institutions recently have developed a heightened awareness of the importance of data recovery and back up in view of the many natural disasters and other world events including

the recent bombing of the World Trade Center in New York City.

The traditional prior art approach at data back up involves taking the processor out of service while back up tapes are made. These tapes are then carried off premises for storage purposes. Should access to the backed up data be required, the proper tape must be located, loaded onto a tape drive, and restored to the host system requiring access to the data. This process is very time consuming and cost intensive, both in maintaining an accurate catalog of the data stored on each individual tape, as well as storing the large number of tapes required to store the large amounts of data required by these institutions. Additionally and most importantly, it often takes 24 hours before a back up tape reaches its' storage destination during which time the back up data is unavailable to the user.

Providers of prior art data storage systems have proposed a method of data mirroring whereby one host CPU or processor writes data to both a primary, as well as a secondary data storage device or system. Such a proposed method, however, overly burdens the host CPU with the task of writing the data to a secondary storage system and thus dramatically impacts and reduces system performance.

Currently, data processing system users often maintain copies of their valuable data on site on either removable storage media, or in a secondary "mirrored" storage device located on or within the same physical confines of the main storage device. Should a disaster such as fire, flood, or inaccessibility to a building occur, however, both the primary as well as the secondary or backed up data will be unavailable to the user.

Accordingly, more data processing system users are requiring the remote storage of back up data.

Accordingly, what is required is a data processing system which automatically and asynchronously with respect to a first host system, generates and maintains a back up or "mirrored" copy of a primary storage device at a location physically remote from the primary storage device, without intervention from, or degrading system performance of the data transfer link between the primary host computer and the primary storage device.

Additionally, today's systems require a significant amount of planning and testing in order to design a data recovery procedure and assign data recovery responsibilities. Typically, a disaster recovery team must travel to the test site carrying a large number of data tapes. The team then loads the data onto disks, makes the required network connections, and then restores the data to the "test" point of failure so processing can begin, such testing may take days or even weeks and always involves significant humans resources in a disaster recovery center or back up site.

#### SUMMARY OF THE INVENTION

This invention features a system which automatically, without intervention from a host computer system, controls storing of primary data received from a primary host computer on a primary data storage system, and additionally controls the copying of the primary data to a secondary data storage system controller which forms part of a secondary data storage system, for providing a back up copy of the primary data on the secondary data storage system which is located in a geographically remote

location from the primary data storage system. In this invention, copying or mirroring of data from a primary data storage system to a secondary data storage system is accomplished without intervention of a primary or secondary host computer and thus, without affecting performance of a primary or secondary host computer system. In the present invention, primary and secondary data storage system controllers are coupled via high speed communication link such as a fiber optic link driven by LED's or laser. At least one of the primary and secondary data storage system controllers coordinates the copying of primary data to the secondary data storage system and at least one of the primary and secondary data storage system controllers maintains at least a list of primary data which is to be copied to the secondary data storage device. Additionally, the secondary data storage system controller provides an indication or acknowledgement to the primary data storage system controller that the primary data to be copied to the secondary data storage system in identical form as secondary data has been received or, in another embodiment, has actually been written to a secondary data storage device.

Accordingly, data may be transferred between the primary and secondary data storage system controllers synchronously, when a primary host computer requests writing of data to a primary data storage device, or asynchronously of the primary host computer requesting the writing of data to the primary data storage system, in which case the remote data copying or mirroring is completely independent of and transparent to the host computer system.

At least one of the primary data storage system controller and the secondary data storage system controller maintains a list of primary data which is to be written to the secondary data storage system. Once the primary data has been at least received or optionally, stored on the secondary data storage system, the secondary data storage system controller provides an indication or acknowledgement of receipt or completed write operation to the primary data storage system. At such time, the primary and/or secondary data storage system controller maintaining the list of primary data to be copied updates this list to reflect that the given primary data has been received by and/or copied to the secondary data storage system. The primary or secondary data storage system controllers and/or the primary and secondary data storage devices may also maintain additional lists in concluding which individual storage locations, such as tracks on a disk drive, are invalid on any given data storage device, which data storage locations are pending a format operation, which data storage device is ready to receive data, and whether or not any of the primary or secondary data storage devices are disabled for write operations.

Thus, an autonomous, host computer independent geographically remote data storage system is maintained providing a system which achieves nearly 100 percent data integrity by assuring that all data is copied to a geographically remote site, and in those cases when a back up copy is not made due to an error of any sort, an indication is stored that the data has not been copied, but instead must be updated at a future date. Such a system is provided which is generally lower in cost and

requiring substantially less manpower and facilities to achieve than the prior art devices.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram illustrating the system with remote data mirroring according to the present invention;

Fig. 2 is a schematic representation of a portion of an index or list maintained by the system of the present invention to determine various features including which data has been copied to a secondary disk; and

Fig. 3 is a schematic representation of an additional list or index maintained by the system of the present invention to keep track of additional items including an invalid data storage device track, device ready status and write disable device status. These and other features and advantages of the present invention will be better understood when read together with the following drawings wherein:

#### DETAILED DESCRIPTION OF THE INVENTION

The present invention features a system which provides a geographically remote mirrored data storage system which contains generally identical information to that stored on a primary data storage system. Utilizing such a system, data recovery after a disaster is nearly instantaneous and requires little, if any, human intervention. Using the present system, the data is retrieved from a remote device through the host data processing system.

The present invention is shown generally at 10, Fig. 1, and includes at site A, which is a first geographic location, a host



computer system 12 as is well known to those skilled in the art. A host computer system 12 is coupled to a first and primary data storage system 14. The host 12 writes data to and reads data from the primary data storage system 14.

5       The primary data storage system 14 includes a primary data storage system controller 16 which receives data from the host 12 over data signal path 18. The primary data storage system controller 16 is also coupled to a storage device 20 which may include a plurality of data storage devices 22a-22c. The storage  
10       devices may include disk drives, optical disks, CD's or other data storage device. The primary system controller 16 is coupled to the storage device 20 by means of data signal path 24.

      The primary data storage system controller 16 includes at least 1 channel adapter (C.A.) 26 which is well known to those  
15       skilled in the art and interfaces with host processing system 12. Data received from the host is typically stored in cache 28 before being transferred through disk adapter (D.A.) 30 over data signal path 24 to the primary storage device 20. The primary  
20       data storage controller 16 also includes a data director 32 which executes 1 or more sets of predetermined micro-code to control data transfer between the host 12, cache memory 28, and the storage device 20. Although the data director is shown as a  
      separate unit, either one of a channel adapter 26 or disk adapter 30 may be operative as a data director, to control the operation  
25       of a given data storage system controller. Such a reconfigurable channel adapter and disk adapter is disclosed in Applicant's co-pending U.S. Patent Application No. 07/587,253 entitled  
      RECONFIGURABLE, MULTI-FUNCTION DISK CONTROLLER of which the

present Application is a continuation in part, and which is fully incorporated herein by reference.

The primary data storage system 14 according to one embodiment of the present invention also includes a service processor 34 coupled to the primary data storage system controller 16, and which provides additional features such as monitoring, repair, service, or status access to the storage system controller 16.

The primary data storage system controller 16 of the present invention also features at least a second disk adapter 36 coupled to the internal bus 38 of the primary data processing system controller 16. The at least second disk adapter 36 is coupled, via a high speed communication link 40 to disk adapter 42 on a secondary data storage system controller 44 of a secondary data storage system 46. Such high speed, .-to-. communication links between the primary and secondary data processing system controllers 16 and 44 include a fiber optic link driven by an LED driver, per IBM escon standard; a fiber optic link driven by a laser driver, and optionally T1 and T3 telecommunication links. Utilizing network connections, the primary and secondary data storage system controller 16 and 44 may be connected to FDDI networks, T1 or T3 based networks and SONET networks.

The secondary data storage system 46 is located a second site geographically removed from the first site. For this Patent Application, geographically removed site means not within the same building as the primary data storage system. There are presently known data processing systems which provide data mirroring to physically different data storage systems. The

systems, however, are generally within the same building. The present invention is directed to providing complete data recovery in case of disaster, such as when a natural disaster such as a flood or a hurricane, or man made disasters such as fires or bombings destroy one physical location, such as one building.

As in the case of the primary data storage system, the secondary data storage system 46 includes, in addition to the secondary data storage system controller 44, a secondary data storage device 48 including a plurality of storage devices 50a -

50c. The plurality of storage devices on the secondary data storage system 46, as well as the primary data storage system 14, may have various volumes and usages such as a primary data storage device 50a which is primary in respect to the attached storage controller 44 and host 52 in the case of the secondary data storage system 46, or primary storage device 22a with respect to the first or primary host 12.

Additionally, each storage device such as storage device 48 may include a secondary storage volume 50b which serves the secondary storage for the primary data stored on the primary volume 22a of the primary data storage system 14. Similarly, the primary data storage system 14 may include a secondary storage volume 22b which stores primary data received and copied from a secondary site and data processing system 46 and host 52.

Additionally, each storage device 20, 48, may include one or more local volumes or storage devices 22c, 50c, which are accessed only by their locally connected data processing systems.

The secondary storage system controller 44 of the present invention also includes at least a first channel adapter 54 which

may receive data from an optionally connected secondary host 52 or an optionally connected hotsite host or CPU 56. Optionally, the primary host 12 may include a data signal path 58 directly into the channel adapter 54 of the secondary data storage system 46, while the optional secondary host 52 may include an optional data path 60 into the channel adapter 26 of the primary data storage system 14. Although the secondary host 52 illustrated in the Fig. is not required for remote data mirroring as described in the present invention, such a host would be required for data retrieval if both the primary host 12 as well as the primary data storage system 14 are rendered inoperative. Similarly, a hotsite host or CPU 56 may optionally be provided at a third geographically remote site to access the data stored in the secondary data storage system 46.

The high speed link 40 between the primary and secondary data storage systems 14 and 46 is designed such that multiple links between the primary and secondary storage system may be maintained for enhanced ability of data and increased system performance. The number of links is variable and may be field upgradeable. Additionally, service processor 34 of the primary data storage system 14 and service processor 62 of the secondary data storage system 46 may also be coupled to provide for remote system configuration, remote software programming and to provide a host base point of control of the secondary data storage system.

The secondary data storage system controller 44 also includes cache memory 64 which receives data from channel adapter 54 and disk adapter 42, as well as disk adapter 66 which controls

writing data to and from secondary storage device 48. Also provided is a data director 68 which controls data transfer over communication bus 70 to which all the elements of the secondary data storage system controller are coupled.

5        'An additional feature of the present invention is the ability to dynamically reconfigure channel adapters as disk adapters and disk adapters as channel adapters, as described in Applicant's co-pending U.S. Patent Application No. 07/587,247 entitled DYNAMICALLY RECONFIGURABLE DATA STORAGE SYSTEM of which  
10       the present Application is a continuation in part, and which is fully incorporated herein by reference.

      The primary and secondary data storage systems may optionally be connected by means of currently available, off the shelf channel extender equipment using bus and tag or escon  
15       interfaces.

      A present invention is designed to provide the copying of data from a primary data storage system to a physically remote secondary data storage system transparent to the user, and external from any influence of the primary host which is coupled  
20       to the primary data storage system. The present invention is designed to operate in at least two modes, the first being a real-time mode wherein the primary and secondary storage systems must guarantee that the data exists and is stored in 2 physically separate data storage units before i/o completion. That is,  
25       before channel end and device end is returned to the host. Alternatively, the present invention is designed to operate in .-in- time mode wherein the data is copied to the remote or secondary data storage system asynchronously from the time when

the primary or local data processing system returns the i/o completion signal (channel end and device end) to the primary host systems. This eliminates any performance penalty if the communication link between the primary and secondary data storage systems is too slow, but creates the additional needs to manage the situation where data is not identical or in "sync" between the primary and secondary data storage systems.

Thus, in the real time mode, the primary data storage system automatically controls the duplication or copying of data to the secondary data storage system controller transparently to the primary host computer. Only after data is safely stored in both the primary and secondary data storage system, as detected by an acknowledgement from the secondary storage system to the primary storage system, does the primary data storage system acknowledge to the primary host computer that the data is synchronized. Should a disaster or facility outage occur at the primary data storage system site, the user will simply need to initialize the application program in the secondary data storage system utilizing a local host (52) or a commercial hotsite CPU or host 56.

The link between the primary and secondary storage system controllers 14 and 46 may be maintained in a uni-directional mode wherein the primary data storage system controller monitors and controls data copying or mirroring. Alternatively, a bi-directional implementation is disclosed in the present invention wherein either controller can duplicate data to the other controller, transparently to the host computer. Should a disaster or facilities outage occur, recovery can be automatic

with no human intervention since the operational host computer already has an active path (40, 58, 60) to the data through its' local controller. While offering uninterrupted recovery, performance will be slower than in an uni-direcitonal implementation due to the over head required to manage intercontroller tasks.

In the second point -in- time of operation the primary data storage system transparently duplicates data to the secondary data storage system after the primary data storage system acknowledges to the host computer, via channel end and device end, that the data has been written to the storage device and the i/o operation has been completed. This eliminates the performance impact of data mirroring over long distances. Since primary and secondary data are not synchronized, however, the primary data storage system must maintain a log file of pending data which has yet to be written to the secondary data storage device. Such data may be kept on removable, non-volatile media, in the cache memory of the primary or secondary data storage system controller as will be explained below, or in the service processor 34, 62.

Accordingly, a feature of the present invention is the ability of a data storage system to control the transfer or copying of data from a primary data storage system to the secondary data storage system, independant of and without intervention from one or more host computers. Most imporantly, in order to achieve optimum data mirroring performance, such data mirroring or copying should be performed asynchronously with i/o requests from a host computer. Accordingly, since data will not

be immediately synchronized between the primary and secondary data storage systems, data integrity must be maintained by maintaining index or list of various criteria including a list of data which has not been mirrored or copied, data storage locations for which a reformat operation is pending, a list of invalid data storage device locations or tracks, whether a given device is ready, or whether a device is write-disabled. Information must also be included as to the time of the last operation so that the data may later be synchronized should an error be detected.

A feature of the present invention is that both the primary or secondary data storage systems maintain a table of the validity of data in the other storage system. As disclosed in co-pending U.S. Patent Application No. 07/586,796 entitled A SYSTEM AND METHOD FOR DISK MAPPING AND DATA RETRIEVAL of which the present Application is a continuation in part and which is fully incorporated herein by reference, the present system maintains a list of index, utilizing one or more flag bits, in a hierarchical structure, on each physical and logical data storage device.

In the present invention, however such information is kept on both devices for each individual system as well as the other data storage system. Thus, as illustrated in the partial list or table 100, Fig. 2, each data storage system maintains an indication of write or copy pending 102 of both the primary data (M1) 104, and the secondary data (M2) 106. Similarly, an index is maintained of a pending format change since a disk format change may be accomplished. The format pending bits 108



including a first primary bit 110 and a second secondary bit 112 indicate that a format change has been requested and such change must be made on the disk.

Thus, when a host computer writes data to a primary data  
5 storage system, it sets both the primary and secondary bits 104,  
106 of the write pending bits 102 when data is written to cache.  
For these examples, the M1 bit will be on the primary data  
storage system and the M2 bit on the secondary data storage  
system. When the primary data storage system controller disk  
10 adapter writes the data to the primary data storage device, it  
will reset bit 104 of the write pending indicator bits 102.

CLAIMS

1. A system for automatically providing and maintaining secondary data, on a secondary data storage device, which is a generally identical copy of primary data stored on a primary data storage device, wherein said secondary data storage device is geographically physically remote from said primary data storage device, said system comprising:

a primary host computer located in a first geographic location;

a primary data storage system located in said first geographic location, and coupled to said primary host computer, for storing data to be accessed by at least said primary host computer;

at least a secondary data storage system including a secondary data storage system controller and at least one secondary data storage device, said secondary data storage system located in said second geographical location and coupled to said primary data storage system, said secondary data storage system controller responsive to primary data received from said primary data storage system controller and which is to be copied and stored as secondary data on said secondary data storage system in identical form as secondary data, and for providing to said primary data storage system controller, an acknowledgement that said primary data has been received to be stored and copied as secondary data on said secondary data storage device;

said primary data storage system including at least one primary data storage device, for storing said primary data received from said primary host computer; and

30 a primary data storage system controller, coupled to said primary host computer and to said at least one primary data storage device, for receiving data from said primary host computer, for controlling the storing of said primary data on said at least one primary data storage device, for maintaining a list of said primary data which is to be copied to said secondary data storage device and stored as secondary data, and  
35 for coordinating and controlling, without intervention from said primary host computer, the copying of said primary data to said secondary data storage device, and responsive to said acknowledgement from said secondary data storage system controller of successful copying of said primary data from said  
40 primary data storage device to said secondary data storage system, for updating said maintained list of said primary data which is to be copied to said secondary data storage device to indicate that said primary data has been copied to said secondary data storage device.

2. The system of claim 1, wherein said primary data storage system controller coordinates the copying of said primary data to said secondary data storage device synchronously with said primary host computer.

3. The system of claim 1, wherein said primary data storage system controller coordinates the copying of said primary data to said secondary data storage device asynchronously with said primary host computer.

4. The system of claim 1, wherein said primary data storage system controller and said secondary data storage system controller are coupled by a high speed communication link.

5. The system of claim 1, wherein said secondary data storage system controller provides said acknowledgement after said primary data has been received and stored on said secondary data storage system.

6. The system of claim 1, further including a secondary host computer, located in said second geographic location geographically remote from said first geographic location, and coupled to at least said secondary data storage system, for  
5 storing a second quantity of primary data to be accessed by at least said secondary host computer, and for at least retrieving said secondary data stored on said secondary data storage system and copied from said primary data on said primary data storage system.

7. The system of claim 1, wherein said secondary data storage system controller maintains a list of at least primary data which is to be stored and copied to said secondary data storage device as secondary data.

8. The system of claim 7, wherein said primary and said secondary data storage systems maintain said list in semiconductor memory.

9. The system of claim 7, wherein said primary and said secondary data storage systems maintain said list in said primary and said secondary data storage devices respectively.

10. The system of claim 7, wherein said maintained list include at least a list of data which must be copied from said primary data storage device to said secondary storage device, a list of data storage device storage locations for which a format  
5 command is pending and for which an invalid track exists.

- 1/2 -

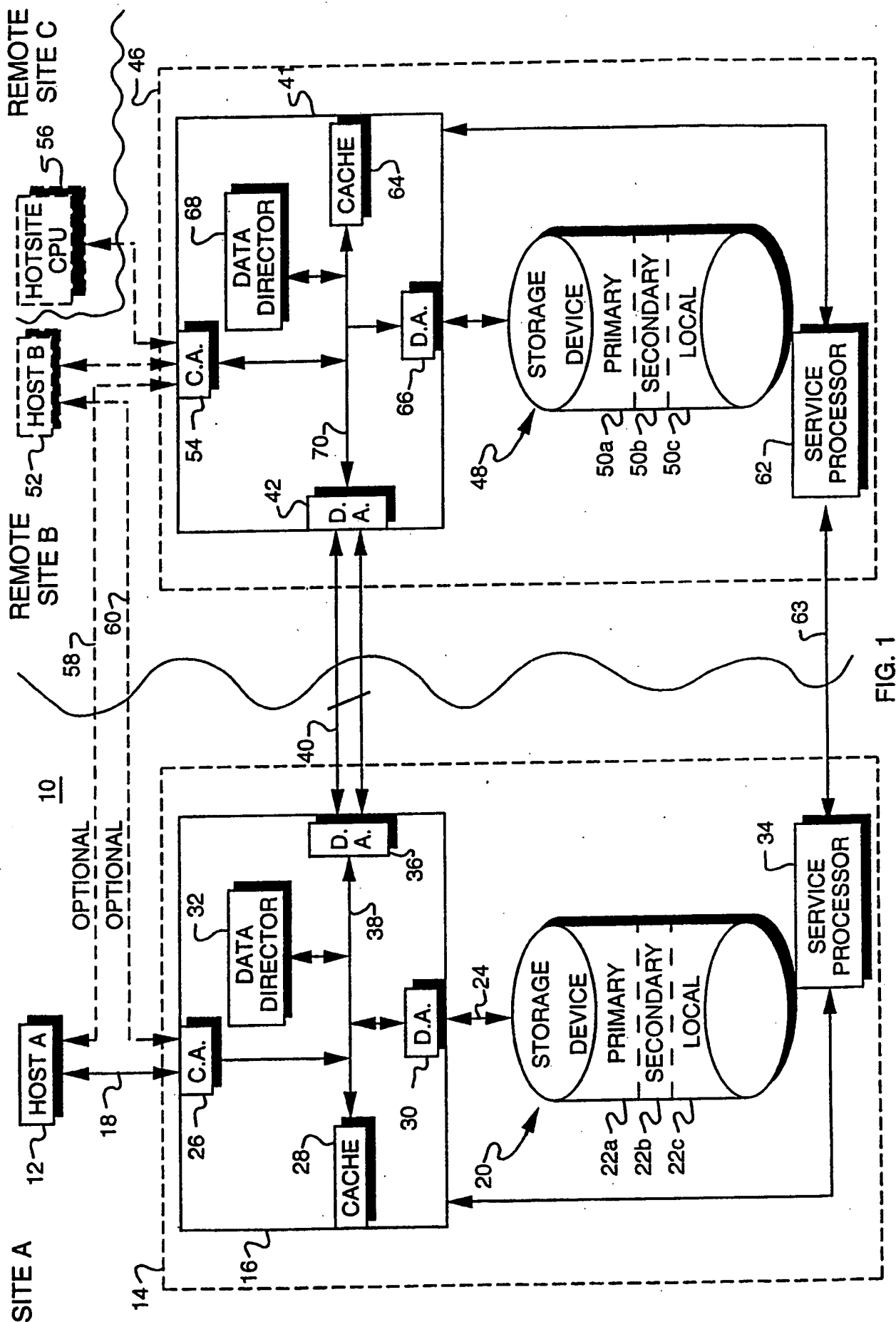


FIG. 1

- 2 / 2 -

100

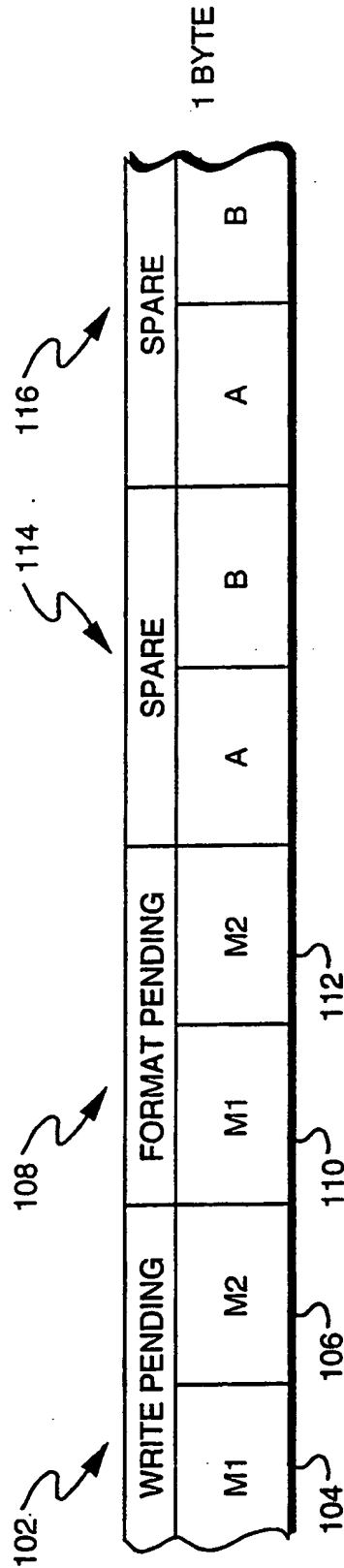


FIG. 2

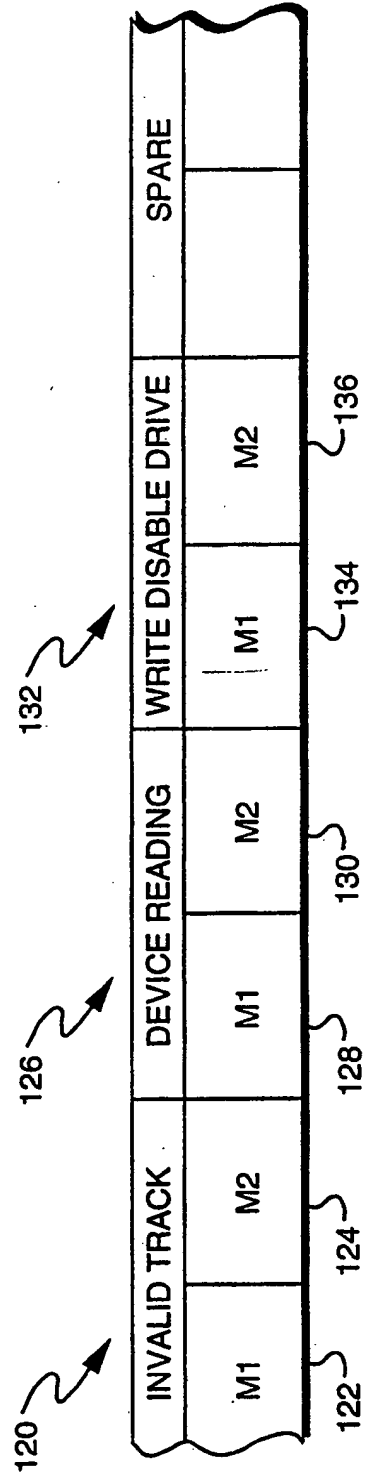


FIG. 3

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US94/04326**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(5) :G06F 12/16

US CL :395/575

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 395/575, 425; 364/200, 900

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
None

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS

searched terms: data copy/transfer/backup, remote, memory, recovery, and failure.

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US, A, 5,146,605 (Beukema et al.) 05 September 1992, fig.s 1-2.	1-10
Y	US, A, 5,051,887 (Berger et al.) 24 September 1991, fig.s 2-3, col. 6 (lines 57-62), col. 8 (lines 32-55), claims 1-7, and corresponding sections in the specification of Berger.	1-10
Y	US, A, 4,755,928 (Johnson et al.) 05 July 1988, background and summary of invention, claim 5, and corresponding sections in the specification of Johnson.	1-10
Y	US, A, 4,710,870 (Blackwell et al.) 05, July 1988, background and summary of invention.	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	*T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be part of particular relevance	*X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z*	document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means		
*P* document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

20 JULY 1994

Date of mailing of the international search report

16 AUG 1994

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-9564

Authorized officer

Matthew M. Kim

Telephone No. (703) 305-3821